Identifying U.S. Merchandise Traders:
Integrating Customs Transactions with Business Administrative Data

by

Fariha Kamal
U.S. Census Bureau

Wei Ouyang
U.S. Census Bureau

CES 20-28 **September, 2020**

Abstract

This paper describes the construction of the Longitudinal Firm Trade Transactions Database (LFTTD) enabling the identification of merchandise traders - exporters and importers - in the U.S. Census Bureau's Business Register (BR). The LFTTD links merchandise export and import transactions from customs declaration forms to the BR beginning in 1992 through the present. We employ a combination of deterministic and probabilistic matching algorithms to assign a unique firm identifier in the BR to a merchandise export or import transaction record. On average, we match 89 percent of export and import values to a firm identifier. In 1992, we match 79 (88) percent of export (import) value; in 2017, we match 92 (96) percent of export (import) value. Trade transactions in year t are matched to years between 1976 and t+1 of the BR. On average, 94 percent of the trade value matches to a firm in year t of the BR. The LFTTD provides the most comprehensive identification of and the foundation for the analysis of goods trading firms in the U.S. economy.

Keyword*:* trade transactions, matching, machine learning

JEL Classification: F00; F10; F14

* Author contact information: Fariha Kamal (corresponding author), fariha.kamal@census.gov; Wei Ouyang, wei.ouyang@census.gov.

We thank Lucia Foster, Nathan Goldschlag, J. Bradford Jensen, C.J. Krizan, Shawn Klimek, Lars Vilhuber and participants at the 2015 Center for Economic Studies seminar series, 2nd BEA-Census Research Workshop, 5th International Conference on Establishment Surveys, 2016 American Economic Association Meetings for valuable comments. We thank Zachary Kroff for excellent research assistance.

1. Introduction

International trade is a major contributor to U.S. gross domestic product (GDP). In 2019, international trade in goods and services accounted for over a quarter of GDP. Goods trade accounts for almost three-quarters of this share.[2] Not only is goods trade an important contributor to aggregate economic growth, but goods traders exhibit a premium over firms participating only in domestic markets: they are larger, pay higher wages, and are more productive (Bernard, Jensen, Redding and Schott, 2018). Participation in global markets provides opportunities for firms to reach new customers and access cheaper inputs. Understanding the impacts of trade on firms, workers, and local economies is central to designing policies aimed at promoting economic growth. However, this requires comprehensive and reliable information on the trading activities of firms. This paper focuses on identifying firms' goods trading activities. We describe the construction of the *Longitudinal Firm Trade Transactions Database* (LFTTD) that permits identification of U.S. merchandise trading firms - both exporters and importers.

The LFTTD combines merchandise export (EXP) and import (IMP) transactions from confidential customs declaration forms with administrative data on the universe of U.S. firms in the non-farm, private sector in the Census Bureau's Business Register (BR). Thus, the LFTTD identifies the U.S. firms associated with goods trade shipments. We build on earlier matching efforts by Bernard, Jensen, and Schott (2009) and, henceforth, referred to as legacy LFTTD. We employ deterministic matching using numeric tax identifiers in the transaction records; and probabilistic matching using business names in the transaction records.[3]

[2] Trade in goods and services in 2019: 4.1 and 1.5 trillion USD, respectively (https://www.census.gov/foreign-trade/Press-Release/2019pr/final_revisions/exh1.txt); GDP in 2019: 21.43 trillion USD (https://www.bea.gov/news/2020/gross-domestic-product-fourth-quarter-and-year-2019-advance-estimate).

[3] Business names are only available for Canadian export transaction records; numeric identifiers are only available for non-Canadian export and all import transactions.

We introduce three major improvements compared to the legacy LFTTD. First, we incorporate income tax information by all reporting units in addition to payroll tax information in the BR. Second, we incorporate historic records from all available years of the BR in order to increase the likelihood of finding a match to a firm identifier. Finally, we introduce probabilistic name matching routines that incorporates state-of-the-art machine learning techniques. Our methodological improvements maximize matches between the trade transactions and the U.S. firm undertaking the shipment, hence, providing the most comprehensive source of information on goods traders in the U.S. economy. All references to LFTTD in this paper, unless otherwise noted, describes the augmented version.[4]

The LFTTD matches an average of 89 percent of value and count of merchandise export transactions to a U.S. firm; and an average of 90 percent of value and 87 percent of count of merchandise import transactions to a U.S. firm. These match rates represent a significant improvement over average match rates in the legacy LFTTD. The legacy LFTTD matches 75 (73) percent of export value (transactions) and 83 (79) percent of import value. The average match rates in the LFTTD mask heterogeneity across years. There is a thirteen (eight) percentage point increase in the value weighted export (import) match rates between 1992 and 2017. This reflects wider availability of numeric identifiers in the trade transactions data over time due to increased automation of the filing process.

The rest of the paper is organized as follows. Section 2 describes the source data. In Section 3, we describe the matching algorithm employed to link merchandise trade transactions to business administrative data and report the match rates in Section 4. The final section provides a discussion of

[4] Qualified researchers on approved projects may access the underlying confidential microdata files described in this paper, for statistical analyses, through the Federal Statistical Research Data Centers (https://www.census.gov/fsrdc).

ongoing research efforts to improve the microdata infrastructure to identify U.S. services trading and multinational firms.

2. Source Data

The U.S. Census Bureau's BR, containing the list of all businesses in the U.S. with paid employees, is the core dataset to which we link the customs transactions data. The merchandise trade transactions are the universe of export and import transactions collected and maintained by the U.S. Customs and Border Protection (CBP) and the U.S. Census Bureau.

2.1 Business Register

The BR covers all U.S. business establishments with paid employees. The core data is sourced from income and payroll tax filings reported to the Internal Revenue Service (IRS) and enhanced with Census Bureau collections to identify the establishments and firms associated with IRS tax identifiers known as employer identification numbers (EIN). The EIN is a nine-digit numeric code identifying a tax entity.[5] The BR contains the particular tax unit identified through tax records including the establishments and firms associated with an EIN. DeSalvo, Limehouse and Klimek (2016) provide detailed description of the sources and functions of the BR with a focus on the BR as a linking tool and bridge to other Census Bureau data.

The BR contains limited information on firms and establishments operating in industries that are outside the scope of the Economic Census.[6] We do not have information about the activity or

[5] See https://www.irs.gov/Businesses/Small-Businesses-&-Self-Employed/Employer-ID-Numbers-EINs for more detail.

[6] Industries outside the scope of the Economic Census include: Agriculture, Forestry and Fishing, Railroads, U.S. Postal Service, Certificated Passenger Air Carriers, Elementary and Secondary Schools, Colleges and Universities, Labor Organizations, Political Organizations, and Religious Organizations. Public administration and governmental entities (NAICS sector 92) are also out of scope with the exception of state-run liquor stores, central reserve depository institutions, federal and federally-sponsored non-depository institutions and hospitals.

location of the establishments associated with employers operating in the out of scope industries and hence cannot determine whether multiple employers fall under common ownership or control of a firm. We only have basic administrative data for these entities. The BR serves as the sampling frame for economic censuses and surveys, a repository of administrative data, and source data for Census public use products including the County Business Patterns and the Business Dynamics Statistics.[7] We use the BR files as described in Lawrence, Stinson, and White (2018) that contain records for establishments with positive payroll in a given year.

2.1 Merchandise Trade Transactions

The U.S. Census Bureau maintains the universe of merchandise export and import transaction records collected by the U.S. CBP starting in 1992. The Center for Economic Studies (CES) at the Census Bureau receives and maintains the confidential, trade transactions files once they have undergone processing required to produce official trade statistics (U.S. Census Bureau, 2014). The EXP and IMP files are described below.

2.1.1 Merchandise Export Transactions

The U.S. CBP collects all export shipments valued above $2,500 to all countries except Canada. The United States substitutes Canadian import statistics for U.S. exports to Canada in accordance with a 1987 Memorandum of Understanding signed by the Census Bureau, U.S. CBP, Canadian Customs, and Statistics Canada.[8] The data exchange only includes U.S. exports destined for Canada but excludes shipments destined for third countries by routes passing through Canada or shipments of certain grains and oilseeds to Canada for storage prior to exportation to a third country.[9]

---

[7] County Business Patterns: https://www.census.gov/programs-surveys/cbp.html; Business Dynamics Statistics: https://www.census.gov/ces/dataproducts/bds/.

[8] This is a reciprocal data exchange where the U.S. provides Canada with U.S. merchandise import shipments to Canada that Canada substitutes for exports to the U.S. (Mozes and Oberg, 2002).

[9] Shipments routed through Canada to a third country must be filed with U.S. CBP.

The U.S. Census Bureau estimates low value (less than $2,500) shipments to all countries except Canada.[10]

Qualified exporters, forwarders or carriers can file export transaction reports, except if destined for Canada, to the U.S. CBP via the Automated Export System (AES).[11] Qualified exporters, forwarders or carriers include businesses or individuals involved in the physical movement of domestic merchandise out of the United States to foreign countries. The merchandise may originate from within the U.S. Customs territory, a U.S. Customs bonded warehouse, or a U.S. Foreign Trade Zone.

Exports of domestic merchandise include commodities which are grown, produced or manufactured in the United States, and commodities of foreign origin which have been transformed in the United States (including U.S. Foreign Trade Zones), or which have been enhanced in value by further manufacture in the United States. Exports of foreign merchandise (re-exports) are also included in the EXP. These consist of commodities of foreign origin which have entered the United States for consumption or in to Customs bonded warehouses or U.S. Foreign Trade Zones, and which, at the time of exportation, are in substantially the same condition as when imported. Within the aforementioned limitations, the merchandise export transactions data are compiled from the universe of official administrative records and therefore are not subject to sampling error.

The EXP data contains an EIN for the entity that ships the exported good to all countries except Canada. Although, majority of exporters file using EINs, an exporter may also file using a social security number (SSN) or foreign entity identification.[12] Additional information collected in the AES

[10] See low value estimation methodology at https://www.census.gov/foreign-trade/aip/lvpaper.html.
[11] The data is collected through electronic filings known as Electronic Export Information (EEI). Before AES, U.S. exporters had to file the Shipper's Export Declaration (SED, see Figure A1), the paper-equivalent of the EEI.
[12] Foreign entity identifiers include foreign passport numbers or Dun & Bradstreet numbers (https://www.census.gov/foreign-trade/aes/documentlibrary/aesparticipantsdata.html).
As of March 24, 2010, social security numbers were no longer permitted as an acceptable form of identification. See https://www.census.gov/foreign-trade/regulations/ftrfaqs.pdf. Transaction records associated with a SSN are flagged in the data and we exclude all identified SSN cases prior to linking non-Canadian transactions to the Business Register.

allows us to distinguish between EIN and non-EIN identifiers. However, prior to the development of the AES in 2006, we do not have a way to distinguish between the different types of identifiers.[13] Since U.S. exporters to Canada do not have to file export declaration forms, we only receive the name of the U.S. business that undertakes the transaction through the data exchange in lieu of a tax identifier. All export transactions include information on the detailed ten-digit Harmonized System product code, the value and quantity of each transaction, the export destination, date of the transaction, the port of exit, mode of transport, whether the shipment was containerized, the U.S. state of origin, and whether the transaction was undertaken between related parties.[14,15] The number of export transactions range from 14 million in 1992 to 39 million in 2017.

We exclude low value estimates and government transactions from our matching procedure. It is important to note that the EIN associated with an export transaction is not a firm identifier. The EIN is a tax identifier for tax reporting purposes but a multi-unit firm can and often has several different such EINs. Hence, EXP is not the universe of qualified exporting firms but rather the universe of qualified tax units used to file export shipments.

2.1.2 Merchandise Import Transactions

The U.S. CBP collects all import transactions valued over $2,000 of merchandise include commodities of foreign origin as well as goods of domestic origin returned to the United States with no change in condition or after having been processed and/or assembled in other countries. The Census Bureau compiles the transaction information from automated data submitted through U.S.

[13] For more information on the AES, see https://www.census.gov/foreign-trade/aes/documentlibrary/bp/2012AESComplianceBestPracticemanual082012.pdf.
[14] See http://www.census.gov/foreign-trade/aes/documentlibrary/aesparticipantsdata.html for a complete list of data elements collected in the AES. Refer to Figure A1 for the paper-equivalent Shipper's Export Declaration. The EXP files received by CES do not contain the name and address of the U.S. exporter or foreign buyer.
[15] For exports, Foreign Trade Statistics Regulations, 30.7(v), define a related-party transaction as one between a U.S. exporter and a foreign consignee, where either party owns, directly or indirectly, 10 percent or more of the other party.

CBP's Automated Commercial System. Data are also compiled from import entry summary forms, warehouse withdrawal forms and Foreign Trade Zone documents as required by law to be filed with the U.S. CBP. The U.S. receives Canadian data on exports of natural gas and electricity to the United States provided by Statistics Canada. The Census Bureau maintains the universe of merchandise import transaction records beginning in 1992. Low value (less than $2,000) shipments to all countries are estimated by the U.S. Census Bureau.

For statistical purposes, imports are classified by the type of transaction: (i) merchandise entered for immediate consumption; (ii) merchandise withdrawn for consumption from Customs bonded warehouses, and U.S. Foreign Trade Zones; and (iii) merchandise entered into Customs bonded warehouses and U.S. Foreign Trade Zones from foreign countries.[16] All import transactions include information on a detailed ten-digit Harmonized System product code, the value and quantity of each transaction, the origin country, the exporting country, date of the transaction, the port of entry, mode of transport, whether the shipment was containerized, the U.S. state of importation, and whether the transaction was undertaken between related parties.[17,18] The number of import transactions range from 16 million in 1992 to 86 million in 2017.

We exclude low value estimates, natural gas, and government transactions from our matching procedure. The importer of record, the entity that receives the shipment, is required to file an import transaction using an employer identification number (EIN), social security number (SSN), or CBP-

[16] The variable "type" in IMP identifies these three categories of transactions. General imports are classified as types "1" and "3"; consumption imports are classified as types "1" and "2".
[17] Refer to Figure A2 for the paper-equivalent of the entry summary, Form 7501. It shows all data elements collected via the CBP's electronic import filing system - Automated Broker Interface. The IMP files received by CES do not contain the name and address of the U.S. importer or foreign supplier. However, it contains item 13, a unique alphanumeric identifier for the foreign supplier (Kamal and Monarch, 2018).
[18] For imports, 19 CFR §152.102(g)) defines related persons as (i) members of the same family, (ii) shared officers or directors, (iii) partners, (iv) employers and employees, and (v) a party having a 5% controlling interest in the other.

assigned number.[19] The EIN is a tax identifier for tax reporting purposes and a firm can and often has several different such EINs. Hence, IMP is not the universe of qualified importing firms but rather the universe of qualified tax units used to file import shipments.

We leverage the information contained in the EXP and IMP to identify all possible U.S. exporting and importing firms, respectively, in the Census Bureau's BR. However, the resulting dataset - LFTTD - does not represent the universe of U.S. exporting or importing firms for three main reasons. First, the BR only contains the set of EINs used for income and payroll tax filings associated with a firm. It is possible that a firm uses a different set of EINs to file merchandise trade shipments. Therefore, EINs in the transaction records need not be in the universe of payroll active EINs known to the Census Bureau. Second, firms may file trade shipments using an identifier other than the EIN (or name) such as a foreign identifier or SSN. Finally, a small but non-negligible share of export and import shipments are missing the identifier information altogether. Bernard, Jensen, and Schott (2009) document that 7 to 10 (3 to 5) percent of export (import) transactions and value have blank identifier fields.

3. Linking Methodology: Trade Transactions and Business Register

We link non-Canadian export and all import transactions to firms in the BR using the EIN when available. We use business name to link Canadian export transactions to firms in the BR.

3.1 Non-Canadian Export and Import Transactions Matching

EINs in the trade transaction records can be linked to the BR to obtain the firm identifier associated with the EIN. Table 1 shows that for shipments with non-missing identifiers, on average,

[19] If a business does not have a EIN or SSN, it can prepare import documentation using CBP assigned numbers using Form 5106 (https://www.cbp.gov/trade/programs-administration/entry-summary/cbp-form-5106). CBP assigned numbers are of the format YYDDMM-NNNNN.

about 85 percent of all non-Canadian export and 85 percent of all import transactions contain an EIN.[20] However, there is substantial heterogeneity across years in the availability of EINs in the export transactions records. For merchandise exports, about 60 percent of all non-Canadian export transactions contained an EIN from 1992 through 1999 with a sharp increase in 2000. Electronic filings were introduced in July, 1995 and became mandatory in 2008 (Federal Register, June 2, 2008).[21] We conjecture that a shift towards electronic filing improved the availability of the EIN variable. For merchandise imports, there is less variation over time. Almost 80 percent or more transactions contain an EIN across years.

All years of the import transaction records contain EINs for the importer of record. Beginning in 2007, it also includes the EIN of the ultimate consignee. The importer of record is liable for payment of all duties and meeting the legal requirements for importation while the ultimate consignee is the actual owner of the merchandise.[22] Given our focus on identifying the U.S. firm that imports, beginning in 2007, we give preference to links made using the ultimate consignee EIN. Table 2 shows that for the majority, 80 percent, of import value and transactions, the EIN is the same for the importer of record and the ultimate consignee.

The EIN in year *t* of filing the merchandise trade transaction might be reported in a different year in the BR. Therefore, we match to multiple years of the BR to improve the likelihood of

[20] Traders may file their EINs in a number of non-standard formats. For example, approximately 40% of export transactions are associated with EINs that are 9-digits but an even higher share, 46% are associated with 11-digits; 12% are 12-digits; and the remaining contain EINs with 8-digits or less. We process EINs with more than nine digits as follows. We create all possible 9-digit combinations of the EIN starting from the left and moving right. For instance, for 12-digit EINs we will create four possible EINs derived from the first nine digits, the second nine digits and so on until no further 9-digit combinations are possible. This process may generate multiple matches such that an export transaction may link to multiple firms. We select a unique match within a year, by choosing the first 9 subset if there is a match, followed by the second 9, and so on.

[21] All federal register notices pertaining to collection of merchandise trade data can be found at https://www.census.gov/foreign-trade/regulations/fedregnotices/index.html.

[22] See Form 7501 for detailed definitions at https://www.cbp.gov/trade/programs-administration/entry-summary/cbp-form-7501.

identifying the firm in a trade transaction. We employ an iterative matching algorithm that proceeds in three main steps. First, we create a bridge file of EINs and associated firm identifiers utilizing a window of years - the current year (*t*), successive year (*t+1*), and preceding year (*t-1*) - of payroll EINs from the BR. Next, we match the residual transaction EINs from the previous step to a second three-year bridge file that links income EINs to a firm identifier.[23] These EINs do not contain payroll or employment information but might provide a link to a firm identifier. Finally, we match the residuals from the second step to a set of historic records, that is, BR files spanning 1976 through (*t-2*).[24]

The window matching may generate multiple matches where an EIN links to a different firm in each window year. If this occurs, we retain the firm match in the current year, followed by the forward year and finally the prior year closest to year *t*. It is possible, but rare, that an EIN matches to multiple firm identifiers within the same year. In such cases, we retain the multi-unit firm following the prevalent finding in the trade literature that exporting firms tend to be larger than non-exporting firms. In cases where an EIN matches to multiple multi-unit firms, we retain the firm with the largest employment. Thus, in our matching algorithm, it is possible to match a trade transaction to a firm that does not have an active EIN in the year that the trade transaction was filed. At the end of the matching process for non-Canadian export and import transaction records, we obtain a crosswalk of matches between a transaction EIN and a unique firm identifier in the BR.

3.2 Canadian Export Transactions Matching

The Canadian export transaction records do not contain an EIN but only the name of the U.S. business. We match the business name to business names in the BR in order to attach a firm

[23] For the most recent year of the trade transactions, the three-year BR bridge files are by construction for two years only. For example, EINs from the 2017 export transactions data are matched to EINs from the 2016 and 2017 BR since the 2018 BR was not yet available at the time of constructing the 2017 LFTTD.

[24] We perform historic matching to the earliest available year of the BR to allow for links to EINs that may have been used by firms in the past.

identifier to the Canadian export transaction. For this exercise, we only consider three years of the BR using both payroll and income EINs around the transaction year due to computational feasibility. Our name matching procedure is iterative and proceeds in four main steps.

1. We first conduct an iterative word match routine that is repeated on three versions of the business name – raw original text from the transaction records, clean name (upcase, add spaces between "&" and "-") and standardized name (clean name and transform words such as "DIVI" to "DIV"). In the first step of the word match routine, we directly match the names against the BR. In the second step, the name string is split into five words (removing words such as 'INC', 'CO', 'CORP', and the like) and then recombined into a string and matched against the BR.[25] The third step removes the character "&" from the name while in the fourth step numbers are further removed from the name. The fifth step is identical to the second step except that we use a maximum of three words only.
2. Once all five steps are completed the remaining unmatched names from the export transactions are subject to a fuzzy name matching procedure. As before, we first normalize the raw names in the two files – EXP and BR - to correct for common misspellings.[26] We match to the BR using SAS DQMATCH and a sensitivity of 95.[27] While the DQMATCH function uses a default sensitivity of 85, a higher value results in stricter matches with 95 being the highest possible threshold. We retain matches to a unique firm. Matches to multiple firm identifiers are processed similarly to non-Canadian export and all import transactions that match to multiple firm identifiers. When no match is identified we again match the name

[25] The business names in the export transactions are a maximum of 30 characters and informs our choice of five words.
[26] In addition, we use DQ Standardization in SAS using 'organization'.
[27] DQMATCH is a fuzzy matching tool that recognizes strings that match inexactly but actually represent the same firm name in the context of our study. The DQMATCH function creates match codes for strings based on their characters, position, and sensitivity. Names sharing same match codes are identified as matches. See http://support.sas.com/resources/papers/proceedings14/1850-2014.pdf for details.

strings now by each of its constituent words in turn separately. Matches up to a minimum of two words are retained, starting with the maximum number of possible words matched.

3. Third, for the remaining names that we were not able to link to a firm in the BR, we check if they were matched to a firm in the Exporter Database (EDB). EDB is developed and maintained by the Census Bureau to produce the "Profile of U.S. Importing and Exporting Companies" report.[28] The EDB differs from the LFTTD primarily due to its reliance on matching to the current year of the BR only and use of different name matching routines that does not include machine learning.[29] However, the EDB incorporates high quality clerical matches for Canadian transactions.[30] Since we do not conduct clerical matches in our matching algorithm, we rely on the EDB to capture additional matches that may be missed through the automated process. The EDB is available starting in 1999. We link years of unmatched Canadian transaction records prior to 1999 to a master file of Canadian transaction records matched to a firm in all available years of the EDB.[31] For years after 1999, we match to the contemporaneous year of the EDB. We also incorporate clerical matches performed to create the legacy LFTTD.
4. Finally, we employ a machine learning (ML) algorithm on the residuals of the unmatched names from the previous three name matching routines.[32] We describe the process below.

3.2.1 Machine Learning Matching

[28] See https://www.census.gov/foreign-trade/statistics/press-release/index.html for complete list of publications of the Profile of U.S. Importing and Exporting Companies.
[29] See https://www.census.gov/foreign-trade/Press-Release/edb/2018/explain.pdf for most recent methodology to construct the EDB.
[30] We also retain matches from EDB for non-Canadian export shipments prior to 2008 in cases of paper filings where clerical errors are more likely.
[31] We create a master file of EDB matches in 1999-2016.
[32] We only use a single year of the BR for the ML algorithm.

The ML algorithm proceeds in four parts: computer assisted translation (CAT), locality sensitive hashing, word pair scoring, and machine learning. We describe each in turn below.

- Computer assisted translation: The purpose of this algorithm is to reduce the absolute number of matches. After gathering the business names from EXP and BR, the CAT algorithm converts all characters to uppercase and removes spacing; keeps only valid alpha characters; corrects suffix misspellings; standardizing suffixes ('LIMITED' to 'LTD', 'PARTN' to 'PARTNER', etc.) and runs through the SAS DQ standardize algorithm. DQ Standardize modify text provided in the appropriate case, with insignificant blank spaces and punctuations removed, standardize certain words and abbreviations, and reorders certain words.
- Locality sensitive hash paring: The purpose of this algorithm is to reduce the number of possible matches. Therefore, we group names together that might reasonably match. The algorithm segments the names into individual words and creates hash codes for each individual word. The individual hash algorithms include standardized word; soundex (Fan, 2004); SAS DQ match 95 (Ordowich *et al*, 2012); caverphone (Hood, 2002); double metaphone (Philips, L. (2000); and NYSIIS (Koneru, Pulla, and Varol, 2016). Matches are retained if any of the segmented words phonetically match and have at least one fifth of their characters in common (such as a grouping between "ALAMAC" and "ALAMACK CORP"). The algorithm results in multiple groupings of possible matches.
- Word pair scoring: This algorithm generates scores to be used as features in the machine learning algorithm. These scoring algorithms take two input strings and output a similarity/distance score. There are twenty-seven different score algorithms in total: three from SAS (Roesch 2012), three custom written, and others in python (Christen, 2008). Some distance metrics include ngram and Jaro-Winkler score (Cohen, Ravikumar, and Fienberg,

2003) among others. Cuffe and Goldschlag (2018) also use a combination of scores in a ML setting.

- Machine Learning: The ensemble algorithm used is stacked generalization (Wolpert, 1992). Stacking is an ensemble method that uses a new model to learn how to best combine the predictions from multiple models trained on the data. The base ML classification learners used are logistic regression (Hilbe, 2009), gini decision tree (Therneau and Atkinson, 1997), and conditional inference tree (Hothorn, Hornik, and Zeileis, 2015). The next stacked generalization level uses the base learners and the base features which best predicts a match as their new input into using the same three ML algorithms. The final layer computes the score by averaging probabilities of the learners of the previous layer. The highest probability match is kept and if there are multiple candidates within that, we select the firm with highest employment.

The window matching may generate multiple matches where a name links to a different firm in each window year. If this occurs, we retain the firm match in the current year, followed by the forward year and finally the prior year closest to year *t*. It is also possible that a name matches to multiple firm identifiers within the same year. In such cases, we retain the multi-unit firm following the prevalent finding in the trade literature that exporting firms tend to be larger than non-exporting firms. In cases where a name matches to multiple multi-unit firms, we retain the firm with the largest employment. Thus, in our matching algorithm, it is possible to match a trade transaction to a firm that is active in a year other than the year trade transaction was filed. At the end of the matching process

for Canadian export and import transaction records, we obtain a crosswalk of matches between a transaction name and a unique firm identifier in the BR.[33]

4. Longitudinal Firm Trade Transactions Database (LFTTD)

The final product of our matching exercise between export and import transaction records and the BR is a set of augmented EXP and IMP files for each year beginning in 1992 through the most recent available year. The LFTTD is not the universe of goods trading firms in the U.S. – it is the set of *known* goods exporting and importing firms. It is important to note that the LFTTD permits identification of firm- and not establishment-level trade flows. It is straightforward to assign trade flows from the LFTTD to single-unit firms. However, for multi-unit firms, especially, whose activities span multiple industries, allocating trade value to their establishments is complicated due to limited availability of product-level information on establishments' output and inputs.[34]

The LFTTD is typically produced with a two-year lag.[35] For example, the 2017 LFTTD is available in 2019. This is dictated by the availability of the final BR files used to create the LFTTD. Three additional variables are added to EXP and IMP files as shown in Table A1. We provide a 10-digit firm identifier that is common across other Census Bureau data sets; the year of the BR from

[33] We note here that our selection criteria introduces bias in the set of identified exporters where we are more likely to select a large firm. We are continuing to investigate algorithms to validate and improve our selection criteria.

[34] The Census of Manufactures (CMF) collects the value of goods exports at the establishment-level, although, not differentiated by products or destination countries. The CMF also collects information on the products produced and used as inputs at the plant. Boehm, Flaaen, and Pandalai-Nayar (2019) describe challenges in classifying firm-level trade into intermediate and final goods. Boehm, Flaaen, Pandalai-Nayar, and Schlupp (2020) describe efforts to allocate export flows from the LFTTD to a firm's manufacturing establishments. Goods export information is not collected for non-manufacturing establishments except in the wholesale trade sector. Moreover, goods import information is not collected for any sector at the establishment-level. See https://www.census.gov/programs-surveys/economic-census/technical-documentation/questionnaires.html for survey methodology and data items collected in the Economic Census.

[35] The export and import transactions data are available with a one-year lag. Qualified researchers requesting the LFTTD should only request the EXP and IMP if there is a research need to employ more contemporaneous data with the caveat that it will not include firm identifiers.

which we obtained the firm identifier; and a set of flags indicating the matching algorithm used to obtain the firm identifier in the BR.[36]

Table A2 lists all available versions of the LFTTD. The LFTTD described in this paper is version D201701 and the legacy LFTTD is version A201101. The first letter denotes any major changes to the matching algorithm; the next four numbers denote the last year in the series; and the final two letters denote any minor changes to the matching algorithm. The LFTTD is a series of files, separately for export and import transactions, by year such that each version of the LFTTD contains two files for each year beginning in 1992 and ending in the year indicated in the version number.[37]

4.1 Matching Results

We match to prior and forward years of the BR relative to the year of the trade transaction to improve the likelihood of identifying the EIN or name in the transaction to an active firm. Most of the matches are obtained in the year of the transaction as shown in Table 3. For export and import transactions on average across 1992-2017, 94 percent of the trade value are linked to a firm identifier in year *t*; 4 percent from years prior to *t*; and 1 percent from the forward year. The shares by count of transactions is very similar. This average pattern (by value and count) is remarkably stable across years as confirmed in unreported results.

Window matching may result in multiple matches. Moreover, name matching often results in matches to multiple firm identifiers. Table 4 displays the share of value and count of trade transactions that are linked to a unique firm identifier (one-to-one match); those that are linked to multiple firm

[36] Although the majority of firm identifiers are obtained from year *t* of the BR, it is possible that some firm identifiers in a given year of the LFTTD cannot be linked to other Census Bureau datasets in the same year. This is also a feature of the legacy LFTTD that relies on three-year window matching (Bernard, Jensen, Schott, 2009: Table 14A.1).

[37] Versions using the same matching methodology will contain identical files except for the most current year and the year prior. The prior year may contain different matches since it is updated with *t+1* of the BR. For example, B201501 and B201401 will contain identical sets of files for years 1992 through 2013. B201501 contains an additional year, 2015, but the 2014 files may also be different since it has been updated using 2015 BR which was not available at the time B201401 was constructed.

identifiers (one-to-many match); and in case of name matching, the share linked to a unique firm identifier using clerical matching.

For matches using EIN in the non-Canadian export and all import transactions, as shown in Panel A, 92 (91) percent of trade value (count of transactions) is linked to a unique firm identifier. Only, 8 (9) percent of trade value (count of transactions) is linked to multiple firm identifiers and require a selection algorithm, as described in Section 3, to select a single firm identifier.

For matches using name in the Canadian export transactions, as shown in Panel B, a little over half of the value is linked to a unique firm identifier; almost a third link to multiple firm identifiers and we select a unique firm identifier using the aforementioned criteria; and a fifth are linked to a unique firm identifier using clerical matching. The count shares are broadly similar. This demonstrates, unsurprisingly, that we are more likely to link an EIN than business name to a unique firm identifier.

4.1.1 Export Transactions

Figure 1 plots the match rates for export transaction records by value between 1992 and 2017. The dark lines show match rates from the LFTTD and the gray lines shows match rates from the legacy LFTTD which ends in 2011. The overall match rates are given by a solid line and the long (short) dashed line shows match rates for the non-Canadian (Canadian) export transactions. Focusing on the LFTTD, we can see that on average, we match about 89 percent of total export value with higher match rates in 2002 and onwards. We match well over 90 percent of Canadian export transactions in all years. There is larger variation in match rates across years for the non-Canadian export transactions where we achieve match rates in the nineties beginning in 2002. We plot the share of transactions matched in Figure A3 to find very similar patterns.

Table A3 provides detailed match rates by value and count of export transactions, respectively. The statistics are separately shown by the year in which the firm identifier is obtained in the BR and

by the destination - Canada and rest of the world (RoW) - of the export transactions. Excluding records that match to firm identifiers in years other than *t* shown under "All Years", the average match rates by both value and count are about 82 percent. Average match rates by value are comparable for non-Canadian and Canadian export transactions.

4.1.2 Import Transactions

Figure 2 plots the match rates for import transaction records by value between 1992 and 2017. The solid dark (gray) line shows match rates for the LFTTD (legacy LFTTD). Focusing on the LFTTD, on average, we match about 89 percent of total import value with match rates in the mid- to high-eighties until 2006 and consistently in the mid-nineties 2007 onwards. We see a sharp increase in the match rates beginning in 2007. This is the year when the EIN of the ultimate consignee became available. We plot the share of transactions matched in Figure A4 to find similar patterns.

Table A4 provides the match rates for import transaction records by value and count of transactions for the years between 1992 and 2017. Match rates by both value and count are separately computed by the year in which the firm identifier is obtained in the BR. On average, we match about 89 (87) percent of total import value (transactions). Excluding records that match to firm identifiers in years other than *t*, the average match rates are 87 and 84 percent by value and count, respectively.

4.2 Comparison with Legacy LFTTD

We improve upon the linking methodology originally used to develop the legacy LFTTD. We can see this clearly from Figures 1 and 2. Comparing the overall export match rates (solid lines) in Figure 1, we see that the legacy LFTTD match rates are on average 10 percentage points lower. The discrepancy is more than double for Canadian export transactions (short-dashed lines). This can be mainly attributed to the implementation of improved name matching routines including machine learning and leveraging the high-quality EDB clerical matches in the LFTTD. The difference in the

overall import match rates (solid lines) in Figure 2 are much smaller compared to Figure 1. Nonetheless, on average the LFTTD (legacy LFTTD) matches 89 (83) percent of import value with the difference becoming larger in later years.

Tables A5 and A6 display the legacy LFTTD match rates for export and import transaction, respectively, between 1992 and 2011. On average, the legacy LFTTD matches about 75 (73) percent of total value (count) of export transactions as shown in Table A5. Canadian match rates are consistently lower than non-Canadian match rates. On average, 76 (79) percent of the value (count) of non-Canadian export transactions and 64 (61) percent of the value (count) of Canadian export transactions are matched to a firm identifier. On average, the legacy LFTTD matches about 83 (79) percent of total value (count) of import transactions as shown in Table A6. Considering matches to firms in year *t* only in Table A3, we still achieve a higher average match rate by value at 87 percent compared to the legacy LFTTD.

5. Future Research

Our efforts to identify trading firms remain incomplete due to the omission of services traders. Unlike, the merchandise trade transactions data that are consistently collected by a central agency, namely the U.S. CBP and the U.S. Census Bureau, services trade transactions are not subject to the same regulations. Service trade transactions are collected through survey instruments by the Bureau of Economic Analysis (BEA).[38]

In a joint project, the Center for Economic Studies at the U.S. Census Bureau and BEA are pursuing the identification of services traders in the Census Bureau's BR. The joint project also

[38] See https://www.bea.gov/international-surveys-us-international-services-transactions for details on the services trade surveys.

identifies U.S. multinational parent firms and U.S. affiliates of foreign multinational parents in the BR.[39] Firms may be globally engaged on three main dimensions: goods trade, services trade, and multinational activities. Thus, the resulting crosswalks combined with the LFTTD will provide the most complete portrait of globally engaged firms in the U.S: merchandise traders, services traders, and multinational firms.

[39] See https://www.bea.gov/surveys/fdiusurv for details on foreign direct investment in the U.S.; see https://www.bea.gov/surveys/diasurv for details on U.S. direct investment abroad.

References

Bernard, A. B., Jensen, J. B., & Schott, P. K., 2009. A portrait of firms in the U.S. that trade goods. In T. Dunne, J. B. Jensen, and M. J. Roberts (Eds.), Producer Dynamics: New Evidence from Micro Data: 383-410. Chicago, IL: University of Chicago Press.

Bernard, A. B., Jensen, J. B., Redding, S. J., & Schott, P. K., 2018. "Global Firms," *Journal of Economic Literature* 56(2), 565-619.

Boehm, C. E., Flaaen, A., & Pandalai-Nayar, N., 2019. "Input Linkages and the Transmission of Shocks: Firm-Level Evidence from the 2011 Tohoku Earthquake," *Review of Economics and Statistics* 101(1), 60-75.

Boehm, C. E., Flaaen, A., Pandalai-Nayar, N., & Schlupp, J., 2020. "Establishment-level Exporting in the U.S.: New Microdata and Facts," mimeo.

Christen, P., 2008. "Febrl: An Open Source Data Cleaning Deduplication and Record Linkage System With a Graphical User Interface," Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Cohen, W.W., Ravikumar, P., & Fienberg, S.E., 2003. "A Comparison of String Distance Metrics for Name-Matching Tasks," Proceedings of the IJCAI-2003 Workshop on Information Integration on the Web.

Cuffe, J., & Goldschlag, N., 2018. "Squeezing More Out of Your Data: Business Record Linkage with Python," U.S. Census Bureau Center for Economic Studies Paper No. CES-WP-18-46.

DeSalvo, B., Limehouse, F.F., & Klimek, S. D., 2016. "Documenting the Business Register and Related Economic Business Data," U.S. Census Bureau Center for Economic Studies working paper 16-17.

Fan Z., 2004. "Matching character variables by sound: a closer look at Soundex function and Sounds-Like Operator (=*)", mimeo.

Hilbe J. M., 2009. *Logistic Regression Models*, CRC Press, Boca Raton, FL.

Hood D., 2002. "Caverphone: Phonetic Matching Algorithm," Technical Paper CTP060902, University of Otago, New Zealand.

Hothorn T, Hornik K., & Zeileis A., 2015. "ctree: Conditional Inference Trees," https://CRAN.R-project.org/web/packages/partykit/vignettes/ctree.pdf.

Kamal, F., & Monarch, R., 2018. "Identifying Foreign Suppliers in U.S. Import Data," *Review of International Economics* 26(1), 117-139.

Koneru, K., Pulla, V., & Varol, C., 2016. "Performance Evaluation of Phonetic Matching Algorithms on English Words and Street Names-Comparison and Correlation," Proceedings of the 5th International Conference on Data Management Technologies and Applications-Volume 1: DATA, p. 57-64.

Lawrence, J., Stinson, M., & White K. W., 2018. "Upcoming Improvements to the Longitudinal Business Database and the Business Dynamics Database," mimeo.

Mozes, S. & Oberg, D., 2002. "U.S. – Canada Data Exchange: 1990-2201," mimeo. Accessed at https://www.census.gov/foreign-trade/aip/uscanada.pdf.

Ordowich, C., Cheney, D., Youtie, J., Fernandez-Ribas, A., & Shapira, P., 2012. "Evaluating the Impact of MEP Services on Establishment Performance: A Preliminary Empirical Investigation," US Census Bureau Center for Economic Studies Paper No. CES-WP-12-15.

Philips, L., 2000. "The Double Metaphone Search Algorithm," *C/C++ Users Journal*, 18(6), p. 38-43.

Roesch, A., 2012. "Matching Data Using Sounds-Like Operators and SAS® Compare Functions," Proceedings of the 2012 SAS Global Forum.

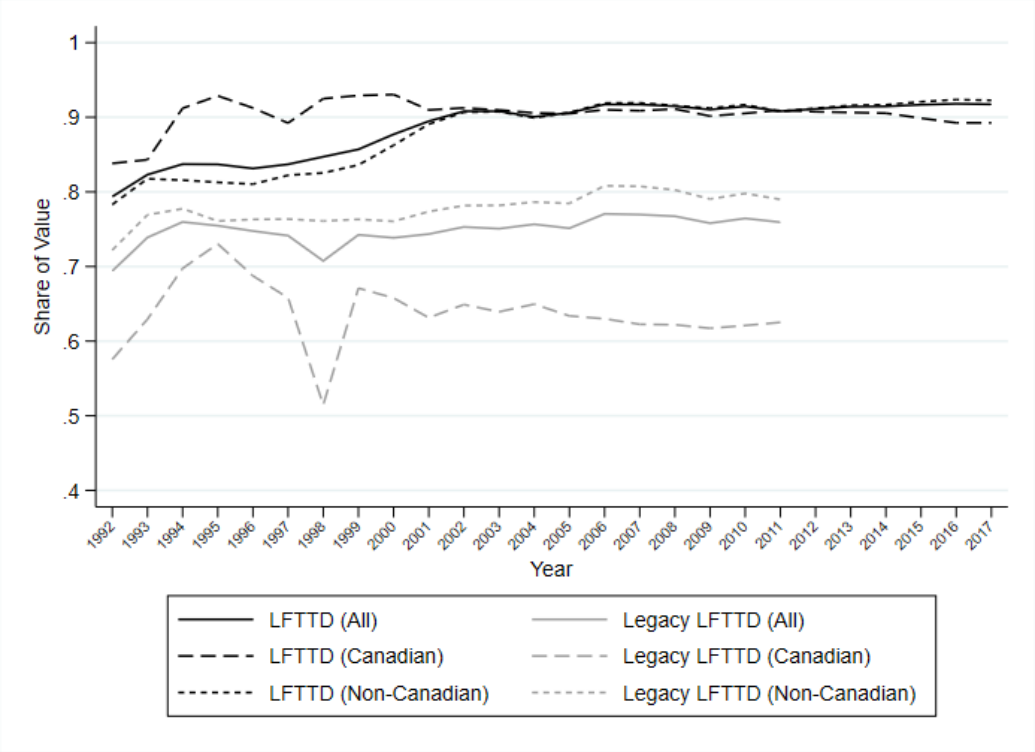
Therneau, T.M., & Atkinson, E.J., 1997. "An Introduction to Recursive Partitioning Using RPART Routines," Technical Report, Mayo Foundation.

U.S. Census Bureau, 2014. "U.S. Merchandise Trade Statistics: A Quality Profile," Accessed at https://www.census.gov/foreign-trade/aip/quality_profile10032014.pdf.

Wolpert, D.H., 1992. "Stacked Generalization," *Neural Networks*, 5(2): p. 241–259.

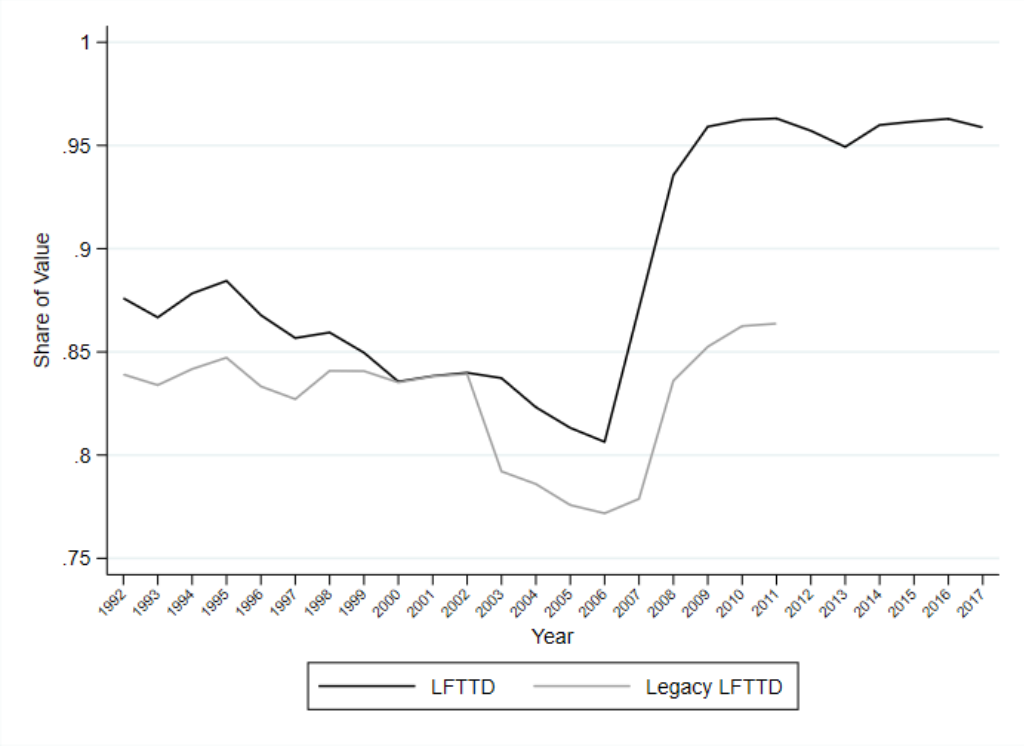
FIGURES

Figure 1: LFTTD Match Rates, Exports (by Value), 1992-2017



Notes: This figure displays the share of export value linked to a firm in the Business Register. Legacy LFTTD refers to the original matching exercise by Bernard, Jensen, and Schott (2009).
Source: Authors' calculations using LFTTD.

Figure 2: LFTTD Match Rates for Imports (by Value)



Notes: This figure displays the share of import value linked to a firm in the Business Register. Legacy LFTTD refers to the original matching exercise by Bernard, Jensen, and Schott (2009).
Source: Authors' calculations using LFTTD.

TABLES

Table 1. Identifiers in Trade Transactions, 1992-2017

	Non-Canadian Export Transactions				Import Transactions			
	By Value		*By Count*		*By Value*		*By Count*	
Year	**EIN**	**Non-EIN**	**EIN**	**Non-EIN**	**EIN**	**Non-EIN**	**EIN**	**Non-EIN**
1992	0.61	0.39	0.52	0.48	0.61	0.39	0.52	0.48
1993	0.60	0.40	0.52	0.48	0.60	0.40	0.52	0.48
1994	0.60	0.40	0.52	0.48	0.60	0.40	0.52	0.48
1995	0.58	0.42	0.52	0.48	0.58	0.42	0.52	0.48
1996	0.56	0.44	0.49	0.51	0.56	0.44	0.49	0.51
1997	0.56	0.44	0.49	0.51	0.56	0.44	0.49	0.51
1998	0.57	0.43	0.51	0.49	0.57	0.43	0.51	0.49
1999	0.60	0.40	0.55	0.45	0.60	0.40	0.55	0.45
2000	0.88	0.12	0.88	0.12	0.88	0.12	0.88	0.12
2001	0.93	0.07	0.92	0.08	0.93	0.07	0.92	0.08
2002	0.95	0.05	0.94	0.06	0.95	0.05	0.94	0.06
2003	0.96	0.04	0.94	0.06	0.96	0.04	0.94	0.06
2004	0.96	0.04	0.95	0.05	0.96	0.04	0.95	0.05
2005	0.97	0.03	0.96	0.04	0.97	0.03	0.96	0.04
2006	0.98	0.02	0.97	0.03	0.98	0.02	0.97	0.03
2007	0.98	0.02	0.97	0.03	0.98	0.02	0.97	0.03
2008	0.98	0.02	0.97	0.03	0.98	0.02	0.97	0.03
2009	0.99	0.01	0.98	0.02	0.99	0.01	0.98	0.02
2010	0.99	0.01	0.99	0.01	0.99	0.01	0.99	0.01
2011	0.99	0.01	0.99	0.01	0.99	0.01	0.99	0.01
2012	0.99	0.01	0.99	0.01	0.99	0.01	0.99	0.01
2013	0.99	0.01	0.99	0.01	0.99	0.01	0.99	0.01
2014	0.99	0.01	0.98	0.02	0.99	0.01	0.98	0.02
2015	0.99	0.01	0.99	0.01	0.99	0.01	0.99	0.01
2016	0.99	0.01	0.99	0.01	0.99	0.01	0.99	0.01
2017	0.99	0.01	0.99	0.01	0.99	0.01	0.99	0.01
Average	**0.85**	**0.15**	**0.83**	**0.17**	**0.85**	**0.15**	**0.83**	**0.17**

Notes: This table displays the share of exporter and importer identifiers that are EINs. Non-EIN identifiers in the non-Canadian export transactions data may include Social Security Numbers (SSN), foreign passport numbers, or Dun & Bradstreet Numbers (DUNS); Non-EIN identifiers in the import transactions data may include Social Security Numbers (SSN) or CBP-assigned numbers.

Source: Authors' calculations using EXP and IMP.

Table 2. Share of Import Value and Transactions by EIN of Importer and Ultimate Consignee, 2007-2017

	Share where EIN of importer and ultimate consignee the same:	
Year	*By Value*	*By Count*
2007	0.85	0.82
2008	0.78	0.81
2009	0.77	0.80
2010	0.77	0.80
2011	0.76	0.80
2012	0.76	0.80
2013	0.77	0.80
2014	0.77	0.80
2015	0.78	0.80
2016	0.79	0.79
2017	0.79	0.79
Average	**0.78**	**0.80**

Notes: Panel A (B) displays the share of exporter (importer) identifiers that are EINs. Non-EIN identifiers in the export transactions data may include Social Security Numbers (SSN), foreign passport numbers, or Dun & Bradstreet Number (DUNS); Non-EIN identifiers in the import transactions data may include Social Security Numbers (SSN) or CBP-assigned numbers.
Source: Authors' calculations using IMP.

Table 3. Year of Match, 1992-2017 Average

Panel A: By Count

	t-2 to 1976	*t-1*	*t*	*t+1*
t	0.04	0.02	0.93	0.01

Panel B: By Value

	t-2 to 1976	*t-1*	*t*	*t+1*
t	0.04	0.02	0.94	0.01

Notes: This table displays the share of trade transactions (Panel A) and the share of trade value (Panel B) accounted for by matches between trade transactions in year *t* and the Business Register in years 1976 through *t+1*.
Source: Authors' calculations using EXP, IMP, and BR.

Table 4. Matching Methods, 1992-2017 Average

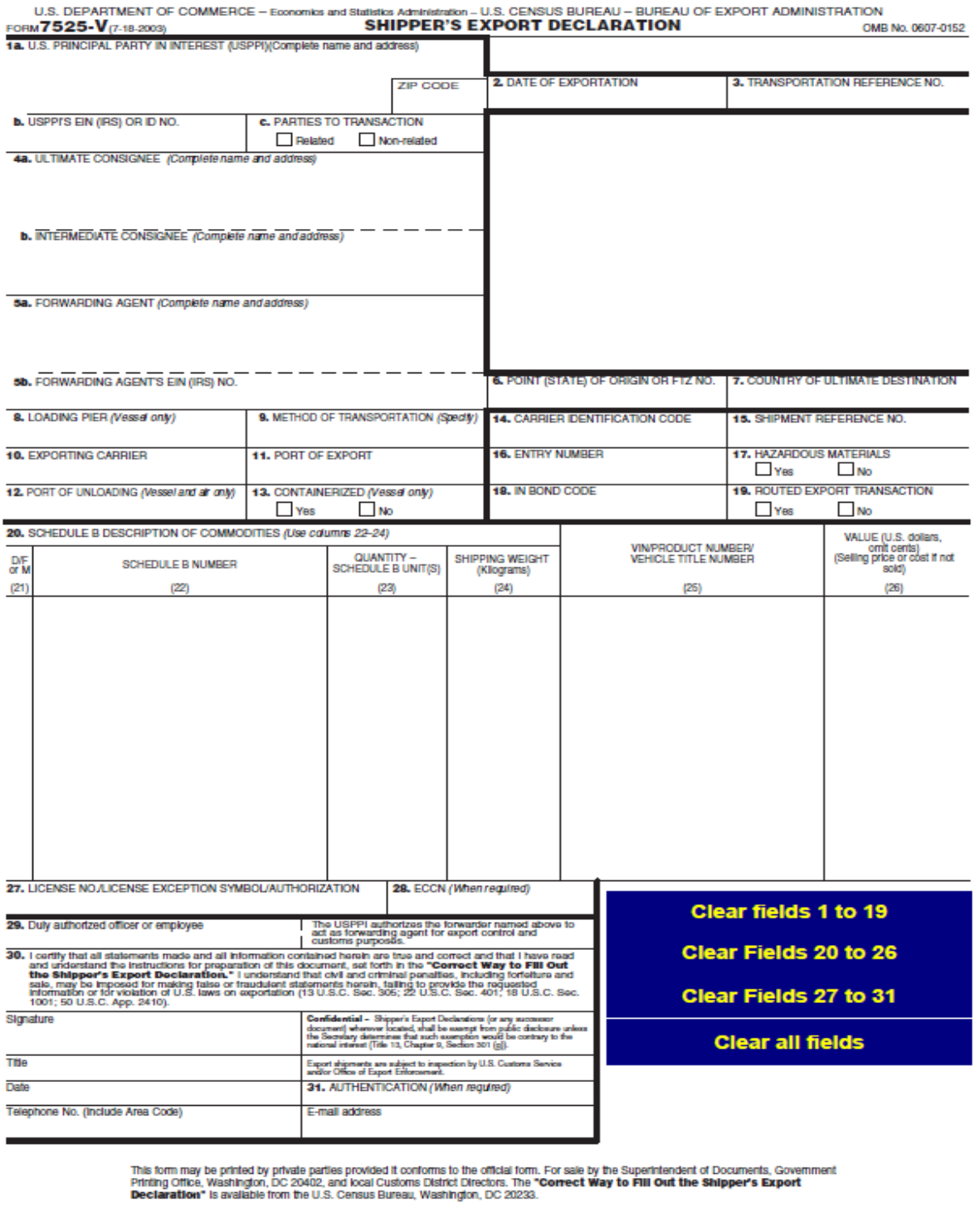
Panel A: EIN Matching		
Match Flag	*By Value*	*By Count*
One-to-one	0.92	0.91
One-to-many	0.08	0.09

Panel B: Name Matching		
Match Flag	*By Value*	*By Count*
One-to-one	0.53	0.54
One-to-many	0.27	0.31
Clerical	0.20	0.14

Notes: This table displays the share of trade value and trade transactions that matched to a unique firm identifier (One-to-one); multiple firm identifiers (One-to-many); and clerically matched to a unique firm identifier (Clerical) in the Business Register. Panel A (B) displays match shares using EINs (names).
Source: Authors' calculations using EXP, IMP, and BR.

Appendix

Figure A1: Form 7525V - Shipper's Export Declaration



U.S. DEPARTMENT OF COMMERCE – Economics and Statistics Administration – U.S. CENSUS BUREAU – BUREAU OF EXPORT ADMINISTRATION

FORM **7525-V** (7-18-2003) **SHIPPER'S EXPORT DECLARATION** OMB No. 0607-0152

1a. U.S. PRINCIPAL PARTY IN INTEREST (USPPI)(Complete name and address)

ZIP CODE

2. DATE OF EXPORTATION

3. TRANSPORTATION REFERENCE NO.

b. USPPI'S EIN (IRS) OR ID NO.

c. PARTIES TO TRANSACTION ☐ Related ☐ Non-related

4a. ULTIMATE CONSIGNEE *(Complete name and address)*

b. INTERMEDIATE CONSIGNEE *(Complete name and address)*

5a. FORWARDING AGENT *(Complete name and address)*

5b. FORWARDING AGENT'S EIN (IRS) NO.

6. POINT (STATE) OF ORIGIN OR FTZ NO.

7. COUNTRY OF ULTIMATE DESTINATION

8. LOADING PIER *(Vessel only)*

9. METHOD OF TRANSPORTATION *(Specify)*

14. CARRIER IDENTIFICATION CODE

15. SHIPMENT REFERENCE NO.

10. EXPORTING CARRIER

11. PORT OF EXPORT

16. ENTRY NUMBER

17. HAZARDOUS MATERIALS ☐ Yes ☐ No

12. PORT OF UNLOADING *(Vessel and air only)*

13. CONTAINERIZED *(Vessel only)* ☐ Yes ☐ No

18. IN BOND CODE

19. ROUTED EXPORT TRANSACTION ☐ Yes ☐ No

20. SCHEDULE B DESCRIPTION OF COMMODITIES *(Use columns 22-24)*

D/F or M (21)	SCHEDULE B NUMBER (22)	QUANTITY – SCHEDULE B UNIT(S) (23)	SHIPPING WEIGHT (Kilograms) (24)	VIN/PRODUCT NUMBER/ VEHICLE TITLE NUMBER (25)	VALUE (U.S. dollars, omit cents) (Selling price or cost if not sold) (26)

27. LICENSE NO./LICENSE EXCEPTION SYMBOL/AUTHORIZATION

28. ECCN *(When required)*

29. Duly authorized officer or employee

The USPPI authorizes the forwarder named above to act as forwarding agent for export control and customs purposes.

30. I certify that all statements made and all information contained herein are true and correct and that I have read and understand the instructions for preparation of this document, set forth in the **"Correct Way to Fill Out the Shipper's Export Declaration."** I understand that civil and criminal penalties, including forfeiture and sale, may be imposed for making false or fraudulent statements herein, failing to provide the requested information or for violation of U.S. laws on exportation (13 U.S.C. Sec. 305; 22 U.S.C. Sec. 401; 18 U.S.C. Sec. 1001; 50 U.S.C. App. 2410).

Signature

Confidential – Shipper's Export Declarations (or any successor document) wherever located, shall be exempt from public disclosure unless the Secretary determines that such exemption would be contrary to the national interest (Title 13, Chapter 9, Section 301 (g)).

Title

Export shipments are subject to inspection by U.S. Customs Service and/or Office of Export Enforcement.

Date

31. AUTHENTICATION *(When required)*

Telephone No. (Include Area Code)

E-mail address

Clear fields 1 to 19

Clear Fields 20 to 26

Clear Fields 27 to 31

Clear all fields

This form may be printed by private parties provided it conforms to the official form. For sale by the Superintendent of Documents, Government Printing Office, Washington, DC 20402, and local Customs District Directors. The **"Correct Way to Fill Out the Shipper's Export Declaration"** is available from the U.S. Census Bureau, Washington, DC 20233.

Source: https://www.ups.com/media/en/shipper_export_dec.pdf.

Figure A2: CBP Form 7501 - Entry Summary

Form Approved OMB No. 1651-0022
EXP. 10-31-2017

DEPARTMENT OF HOMELAND SECURITY
U.S. Customs and Border Protection

ENTRY SUMMARY

1. Filer Code/Entry No.		2. Entry Type	3. Summary Date
4. Surety No.	5. Bond Type	6. Port Code	7. Entry Date

8. Importing Carrier	9. Mode of Transport	10. Country of Origin	11. Import Date
12. B/L or AWB No.	13. Manufacturer ID	14. Exporting Country	15. Export Date

16. I.T. No.	17. I.T. Date	18. Missing Docs	19. Foreign Port of Lading	20. U.S. Port of Unlading
21. Location of Goods/G.O. No.	22. Consignee No.		23. Importer No.	24. Reference No.

25. Ultimate Consignee Name and Address	26. Importer of Record Name and Address
City State Zip	City State Zip

27. Line No.	28. Description of Merchandise: 29. A. HTSUS No. B. ADA/CVD No.	30. A. Grossweight B. Manifest Qty.	31. Net Quantity in HTSUS Units	32. A. Entered Value B. CHGS C. Relationship	33. A. HTSUS Rate B. ADA/CVD Rate C. IRC Rate D. Visa No.	34. Duty and I.R. Tax: Dollars	Cents

Other Fee Summary for Block 39	35. Total Entered Value $ / Total Other Fees $	**CBP USE ONLY**		TOTALS
		A. LIQ CODE	B. Ascertained Duty	37. Duty
		REASON CODE	C. Ascertained Tax	38. Tax
			D. Ascertained Other	39. Other
			E. Ascertained Total	40. Total

36. DECLARATION OF IMPORTER OF RECORD (OWNER OR PURCHASER) OR AUTHORIZED AGENT

I declare that I am the ☐ Importer of record and that the actual owner, purchaser, or consignee for CBP purposes is as shown above, **OR** ☐ owner or purchaser or agent thereof. I further declare that the merchandise ☐ was obtained pursuant to a purchase or agreement to purchase and that the prices set forth in the invoices are true, **OR** ☐ was not obtained pursuant to a purchase or agreement to purchase and the statements in the invoices as to value or price are true to the best of my knowledge and belief. I also declare that the statements in the documents herein filed fully disclose to the best of my knowledge and belief the true prices, values, quantities, rebates, drawback s, fees, commissions, and royalties and are true and correct, and that all goods or services provided to the seller of the merchandise either free or at reduc ed cost are fully disclosed.
I will immediately furnish to the appropriate CBP officer any information showing a different statement of facts.

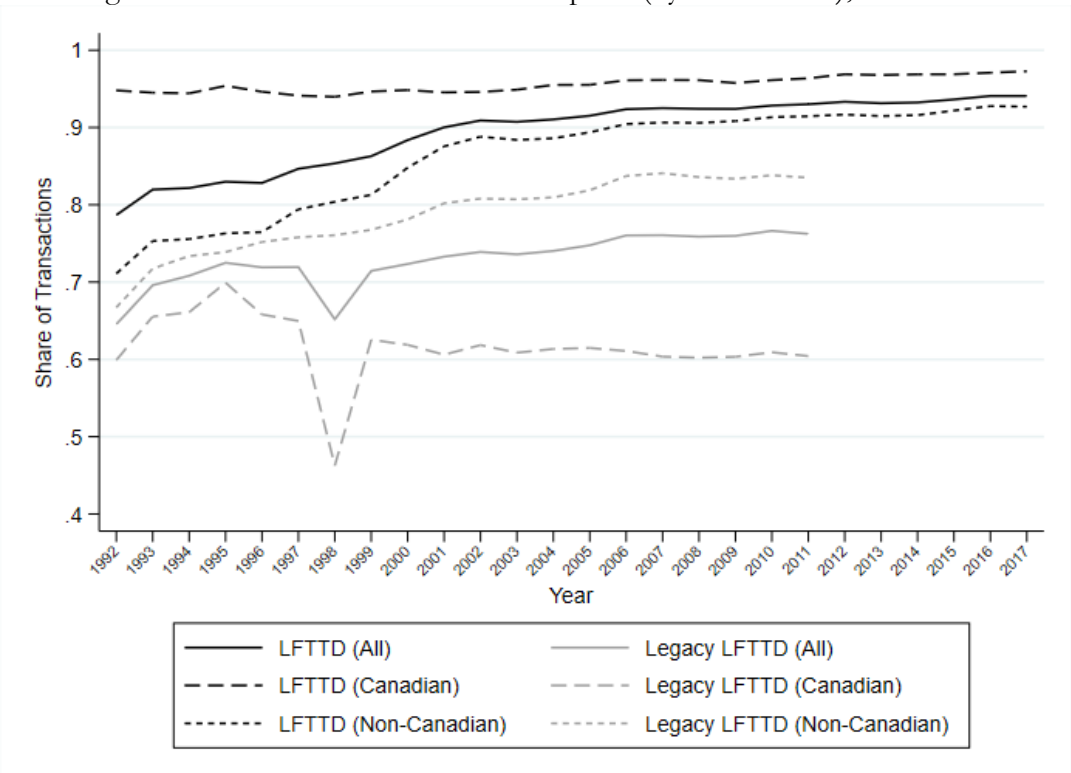
41. DECLARANT NAME TITLE SIGNATURE DATE

42. Broker/Filer Information (Name, address, phone number)	43. Broker/Importer File No.

CBP Form 7501 (06/09)

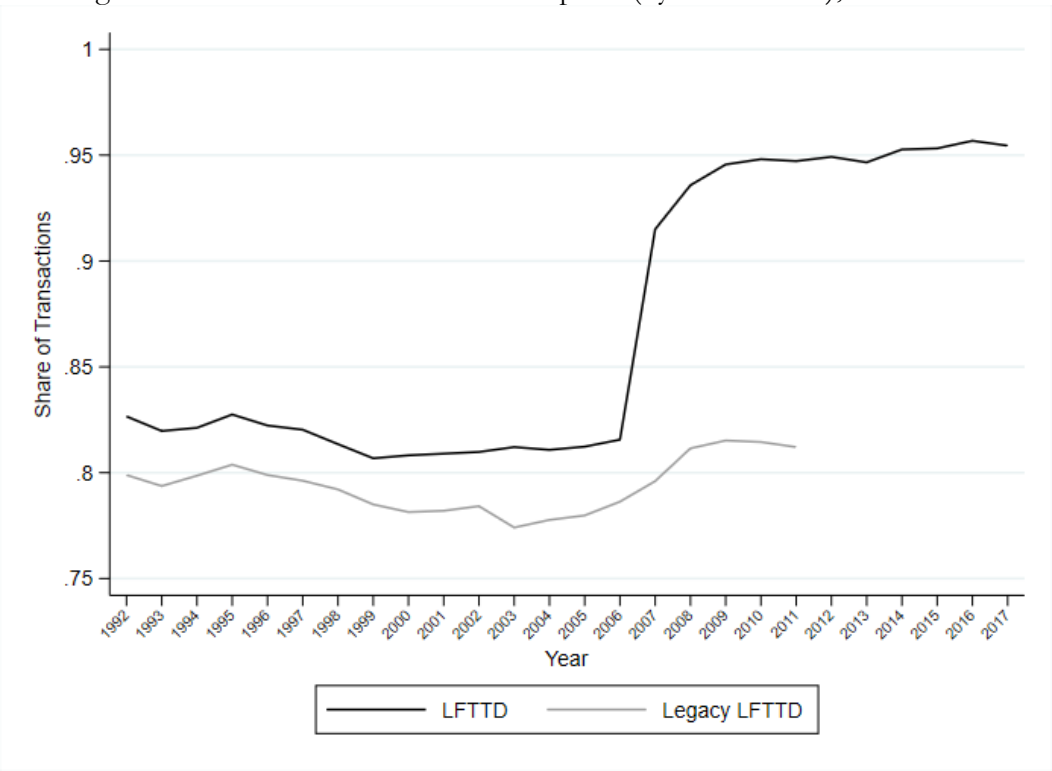
Source: https://www.cbp.gov/sites/default/files/assets/documents/2018-Feb/CBP%20Form%207501.pdf.

Figure A3: LFTTD Match Rates for Exports (by Transaction), 1992-2017



Notes: This figure displays the share of export transactions linked to a firm in the Business Register. Legacy LFTTD refers to the original matching exercise by Bernard, Jensen, and Schott (2009).
Source: Authors' calculations using LFTTD.

Figure A4: LFTTD Match Rates for Imports (by Transaction), 1992-2017



Notes: This figure displays the share of import value linked to a firm in the Business Register. Legacy LFTTD refers to the original matching exercise by Bernard, Jensen, and Schott (2009).
Source: Authors' calculations using LFTTD.

Table A1. LFTTD – Description of Additional Variables

Variable Name	Definition
firmid	10-digit numeric code identifying a firm in the Business Register
match_code	Flags indicating the matching algorithm used to obtain the firmid
firmidyear	Year of the Business Register in which the firmid was obtained

Notes: This table shows the additional variables contained in the LFTTD files.

Table A2. LFTTD – Available Versions

Version	Years
A201101	1992-2011
B201401	1992-2014
B201501	1992-2015
C201601	1992-2016
D201701	1992-2017

Notes: This table shows the currently available versions of the LFTTD. The first letter denotes a major change in matching algorithm where "A" is the legacy LFTTD; the four-digit year indicates the last year in a version; the last two digits indicate minor changes to matching algorithm where "01" is the default.

Table A3. LFTTD Match Rates, Exports, 1992-2017.

	By Value				*By Count*			
	By Year of Match		*By Destination*		*By Year of Match*		*By Destination*	
Year	**All years**	**Year *t***	**RoW**	**Canada**	**All years**	**Year *t***	**RoW**	**Canada**
1992	0.79	0.77	0.78	0.84	0.79	0.76	0.71	0.95
1993	0.82	0.80	0.82	0.84	0.82	0.79	0.75	0.95
1994	0.84	0.81	0.82	0.91	0.82	0.79	0.76	0.94
1995	0.84	0.81	0.81	0.93	0.83	0.80	0.76	0.95
1996	0.83	0.79	0.81	0.91	0.83	0.78	0.76	0.95
1997	0.84	0.80	0.82	0.89	0.85	0.80	0.79	0.94
1998	0.85	0.82	0.83	0.93	0.85	0.82	0.80	0.94
1999	0.86	0.82	0.84	0.93	0.86	0.81	0.81	0.95
2000	0.88	0.84	0.86	0.93	0.88	0.83	0.85	0.95
2001	0.89	0.85	0.89	0.91	0.90	0.85	0.88	0.95
2002	0.91	0.88	0.91	0.91	0.91	0.87	0.89	0.95
2003	0.91	0.88	0.91	0.91	0.91	0.87	0.88	0.95
2004	0.90	0.87	0.90	0.91	0.91	0.87	0.89	0.96
2005	0.91	0.88	0.91	0.90	0.92	0.88	0.89	0.96
2006	0.92	0.89	0.92	0.91	0.92	0.89	0.90	0.96
2007	0.92	0.89	0.92	0.91	0.93	0.89	0.91	0.96
2008	0.91	0.88	0.92	0.91	0.92	0.89	0.91	0.96
2009	0.91	0.88	0.91	0.90	0.92	0.89	0.91	0.96
2010	0.91	0.88	0.92	0.91	0.93	0.89	0.91	0.96
2011	0.91	0.87	0.91	0.91	0.93	0.89	0.91	0.96
2012	0.91	0.88	0.91	0.91	0.93	0.90	0.92	0.97
2013	0.91	0.88	0.92	0.91	0.93	0.90	0.91	0.97
2014	0.91	0.88	0.92	0.91	0.93	0.90	0.92	0.97
2015	0.92	0.88	0.92	0.90	0.94	0.90	0.92	0.97
2016	0.92	0.89	0.92	0.89	0.94	0.90	0.93	0.97
2017	0.92	0.89	0.92	0.89	0.94	0.91	0.93	0.97
Average	**0.89**	**0.85**	**0.88**	**0.90**	**0.89**	**0.86**	**0.86**	**0.96**

Notes: This table displays the share of export value and transactions matched to a unique firm identifier in the Business Register. "Year of Match" refers to the year that the firm identifier was obtained from the Business Register. The denominator under "Year of Match" is the total export value or transaction. "Destination" refers to Canada and rest of the world (RoW). The denominator under "Destination" is total export value or transaction by destination.

Source: Authors' calculations using LFTTD

Table A4. LFTTD Match Rates, Imports, 1992-2017.

	By Value		*By Count*	
Year	**All Years**	**Year *t***	**All Years**	**Year *t***
1992	0.88	0.86	0.83	0.80
1993	0.87	0.85	0.82	0.80
1994	0.88	0.86	0.82	0.80
1995	0.88	0.87	0.83	0.80
1996	0.87	0.85	0.82	0.79
1997	0.86	0.84	0.82	0.79
1998	0.85	0.83	0.81	0.79
1999	0.85	0.83	0.81	0.78
2000	0.84	0.82	0.81	0.78
2001	0.84	0.82	0.81	0.78
2002	0.84	0.82	0.81	0.79
2003	0.84	0.82	0.81	0.79
2004	0.82	0.81	0.81	0.79
2005	0.81	0.80	0.81	0.79
2006	0.81	0.79	0.82	0.79
2007	0.87	0.85	0.92	0.88
2008	0.94	0.91	0.94	0.90
2009	0.96	0.94	0.95	0.91
2010	0.96	0.94	0.95	0.92
2011	0.96	0.94	0.95	0.91
2012	0.96	0.94	0.95	0.92
2013	0.95	0.93	0.95	0.91
2014	0.96	0.94	0.95	0.92
2015	0.96	0.94	0.95	0.92
2016	0.96	0.94	0.96	0.92
2017	0.96	0.93	0.95	0.92
Average	**0.89**	**0.87**	**0.87**	**0.84**

Notes: This table displays the share of import value and transactions matched to a unique firm identifier in the Business Register. "All Years" shows the total match rates; "Year *t*" shows match rates obtained from year *t* of the Business Register.
Source: Authors' calculations using LFTTD.

Table A5. Legacy LFTTD Match Rates, Exports, 1992-2011.

	By Value			*By Count*		
Year	**All**	**RoW**	**Canada**	**All**	**RoW**	**Canada**
1992	0.69	0.72	0.58	0.65	0.67	0.60
1993	0.74	0.77	0.63	0.70	0.72	0.66
1994	0.76	0.78	0.70	0.71	0.73	0.66
1995	0.75	0.76	0.73	0.73	0.74	0.70
1996	0.75	0.76	0.69	0.72	0.75	0.66
1997	0.74	0.76	0.66	0.72	0.76	0.65
1998	0.71	0.76	0.51	0.65	0.76	0.46
1999	0.74	0.76	0.67	0.71	0.77	0.63
2000	0.74	0.76	0.66	0.72	0.78	0.62
2001	0.74	0.77	0.63	0.73	0.80	0.61
2002	0.75	0.78	0.65	0.74	0.81	0.62
2003	0.75	0.78	0.64	0.74	0.81	0.61
2004	0.76	0.79	0.65	0.74	0.81	0.61
2005	0.75	0.78	0.63	0.75	0.82	0.61
2006	0.77	0.81	0.63	0.76	0.84	0.61
2007	0.77	0.81	0.62	0.76	0.84	0.60
2008	0.77	0.80	0.62	0.76	0.84	0.60
2009	0.76	0.79	0.62	0.76	0.83	0.60
2010	0.76	0.80	0.62	0.77	0.84	0.61
2011	0.76	0.79	0.63	0.76	0.84	0.60
Average	**0.75**	**0.78**	**0.64**	**0.73**	**0.79**	**0.62**

Notes: This table displays the share of export value and count of export transactions matched to a unique firm identifier in the Business Register in the legacy LFTTD (Bernard, Jensen, Schott, 2009) in "All"; and separately for exports destined to Canada and the rest of the world (RoW). Legacy LFTTD is only updated through 2011.
Source: Authors' calculations using LFTTD.

Table A6. Legacy LFTTD Match Rates, Imports, 1992-2011.

Year	*By Value*	*By Count*
1992	0.84	0.80
1993	0.83	0.79
1994	0.84	0.80
1995	0.85	0.80
1996	0.83	0.80
1997	0.83	0.80
1998	0.84	0.79
1999	0.84	0.79
2000	0.84	0.78
2001	0.84	0.78
2002	0.84	0.78
2003	0.79	0.77
2004	0.79	0.78
2005	0.78	0.78
2006	0.77	0.79
2007	078	0.80
2008	0.84	0.81
2009	0.85	0.82
2010	0.86	0.81
2011	0.86	0.81
Average	**0.83**	**0.79**

Notes: This table displays the share of import value and count of import transactions matched to a unique firm identifier in the Business Register in the legacy LFTTD (Bernard, Jensen, Schott, 2009). Legacy LFTTD is only updated through 2011.
Source: Authors' calculations using LFTTD.